Patent Office
Research and Development
Reports 2

ADVANCES IN MECHANIZATION OF PATENT SEARCHING

CHEMICAL FIELD

April 11, 1956

FOR OFFICIAL DISTRIBUTION

Prepared by

B. E. Lanham
Deputy Director

J. Leibowitz
H. R. Koller
Patent Research Specialists
Staff Members

Office of Research and Development
U. S. Patent Office

Robert C. Watson
Commissioner of Patents

Sinclair Weeks
Secretary of Commerce

ADVANCES IN MECHANIZATION OF PATENT SEARCHING[1]

HISTORY

A previous experiment in mechanized searching in the U.S. Patent Office, using punched cards for the storage and retrieval of chemical compositions of matter disclosures, has been described (1). That experiment yielded encouraging results with respect to the feasibility of employing machine techniques for the complex types of searches required by patent examiners.

Subsequent to that effort, a Committee advisory to the Secretary of Commerce was formed under the Chairmanship of Dr. Vannevar Bush. The purpose of the Committee was to study the Patent Office search procedures and problems as well as the prior experiment and to report on possible application of mechanized techniques of searching to speed up the handling of applications by providing more rapid and efficient searches.

In accordance with the Committee's recommendations (2) the Patent Office, with the cooperation of the National Bureau of Standards, is now engaged in an effort to put machine searching of compositions of matter on an operational basis. To stimulate and coordinate the program, an Advisory Committee to the Secretary of Commerce has been formed. The Patent Office Annual Report for the Fiscal Year 1955 refers to the general features and status of the previous and present programs (3).

In view of the broad program now contemplated, the mechanized search system must provide much greater flexibility in searching with respect to types of subject matter and variation of search viewpoint than was available according to the earlier mechanization system. This article will discuss briefly some of the general features of the system which have been developed subsequent to the work described in the earlier article.

SOME RELATIONSHIPS TO BE DEALT WITH

In every patent search it is required that all disclosures be located which are of relevance to a particular point of view. The point of view varies with each search and it is desired, in mechanization, to enable machine manipulation of the disclosure data so as to retrieve all information pertinent to any of a number of constantly variable search viewpoints.

To illustrate some of the considerations involved, a searcher may desire to find all documents disclosing a mixture of compounds A and B used as a lubricating composition. Or he may be interested in discovering what uses have been disclosed for a mixture of A and B. In the first situation, the search requirement is satisfied by a disclosure of A mixed with B only where a lubricating function is also found associated with that mixture, whereas the second search requirement is satisfied by a disclosure of admixtures of A and B, regardless of function. As another example, the disclosure of boric acid mixed with alcohol will not satisfy a searcher whose interest is in compounds containing both boron and a hydroxyl group, for while the two named things are disclosed, their relationship is not in accordance with the search requirement.

It is often necessary to relate functions to pertinent materials. Assume, for example, that a mixture of compounds A and B forms an insecticide, although neither of them has that property individually, and that upon the addition of compound C a lubricant is obtained, the insecticidal characteristic not being present in the 3 component mixture. The insecticidal function is associated with each of A and B but not with C, whereas the lubricant function is associated with each of A, B and C. This disclosure may be represented as follows:

$$\underbrace{\overbrace{A \qquad B}^{\text{insecticide}} \qquad C}_{\text{lubricant}}$$

These properties may be considered as "distributed to" or "shared by" each of the materials which contribute to the existence of the properties so that the representation may be transcribed into the following form:

$$\begin{pmatrix} \text{A} \\ \text{D - insecticide} \\ \text{D - lubricant} \end{pmatrix} \quad \begin{pmatrix} \text{B} \\ \text{D - insecticide} \\ \text{D - lubricant} \end{pmatrix}$$

$$\begin{pmatrix} \text{C} \\ \\ \text{D - lubricant} \end{pmatrix}$$

[1] Revision of paper presented before the Division of Chemical Literature, 129th Meeting of American Chemical Society, Dallas, Texas, April 11, 1956.

where the "D" indicates a distributed or shared property. Since the three compounds posses these properties not individually, but only in certain mixture combinations, a distinction must be made between the shared function of each compound which contributes to the function of the entire admixture and the function of the individual ingredient only.

Some materials are disclosed only functionally. Take, for example, a mixture of nitrocellulose, a pigment and a solvent. The latter two terms represent particular materials, but they are described only in terms of their functional characteristics. Consequently, while materials disclosed in this manner can be retrieved in terms of function, they cannot be retrieved in terms of their chemical structure.

A disclosure may show a relationship of alternativeness with or without equivalence. Thus the disclosure that "Compound X may be mixed with either water or gasoline" indicates that water is alternative to gasoline with respect to admixture with compound X. If when water is used, the composition becomes a mouthwash whereas with the use of gasoline an insecticide is formed, the relationship is one of mere alternativeness without equivalence. However, if the same utility is obtained with either water or gasoline mixed with compound X, a condition of equivalence exists between the water and gasoline with respect to the utility involved, in addition to their alternative relationship.

There are many known materials of complex nature which are themselves admixtures of several chemical compounds. Many biological extracts, plant or animal parts or complex reaction mixtures, for example, contain a number of chemical compounds. It is necessary to record each of these compounds and to associate the several compounds together in terms of their being contained in a common mixture having a single source. An example of why this is so is that a disclosure of milk in admixture with sodium carbonate should be available to one interested in casein mixed with sodium carbonate as well as to one interested in a mixture of milk and sodium carbonate.

Several compounds may each be described in terms of different relationships to a particular chemical process, some for example as being starting materials, others as final products. In many disclosures the product of a chemical reaction is not structurally described, and products of this type may be best retrieved in terms of their relationship to the process of making them. For example, product L may be described solely as the product of the oxidation of compound M using compound V as the oxidizing agent.

THE CLASS CONCEPT

The concept of classes is fundamental to the search system. A search for *an* alcohol is a search for any and all things having the characteristics which define the class of alcohols. A search for *a* bactericide is a search for any and all members of the class defined in terms of that function. A class may, of course, be expressed in terms of multiple characteristics, such as the class of bactericidal alcohols.

Since patent searching requires the finding of similarities, all of the class relationships which can be discerned among the pertinent subject matter must be maintained. The machine language used to describe a disclosure should be such as to express the various class relationships. That is, the codes for a material such as aniline must bear sufficient relationship to the codes for benzene to permit retrieval of disclosures of both of these materials in a search for benzene compounds and to permit rejection of one and selection of the other in a search for *an* anilino compound.

The concept of class relationships and the application of the system thereto may be illustrated by Figure 1. An attempt has been made at universality, i. e., without limitation to any specific type of class. Boolean symbology is employed.

Let A, B and C represent classes defined in terms of characteristics A, B and C, respectively. When it is said that Class C is defined in terms of characteristic C it is not meant to say that each member of the class *is* C, but rather that each member of the class has, or may be described in terms of the characteristics of "C-ness". Class A includes as its members things which are members of Class A and no other class and members of Class A and some other class (X representing any other class). Assuming complete identification of all things, so that each thing is uniquely delineated from all other things, Class $A \cap \bar{X}$ would contain only one member. If the characteristic defining Class A is the benzene structure, the compound benzene would be the sole member of Class $A \cap \bar{X}$. A member of Class $A \cap X$ would be any compound containing the benzene configuration together with other groups attached thereto. Therefore, Class A, which is generic to $A \cap X$ and $A \cap \bar{X}$, would include as members benzene itself and all compounds containing the benzene structure somewhere in the molecule. The same principle is similarly shown for two other distinct classes, B and C.

Continuing this analysis, Class $A \cap X$ includes things which are members of Class $A \cap B \cap \bar{X}$, Class $A \cap C \cap \bar{X}$, Class $A \cap B \cap X$ and Class $A \cap C \cap X$.

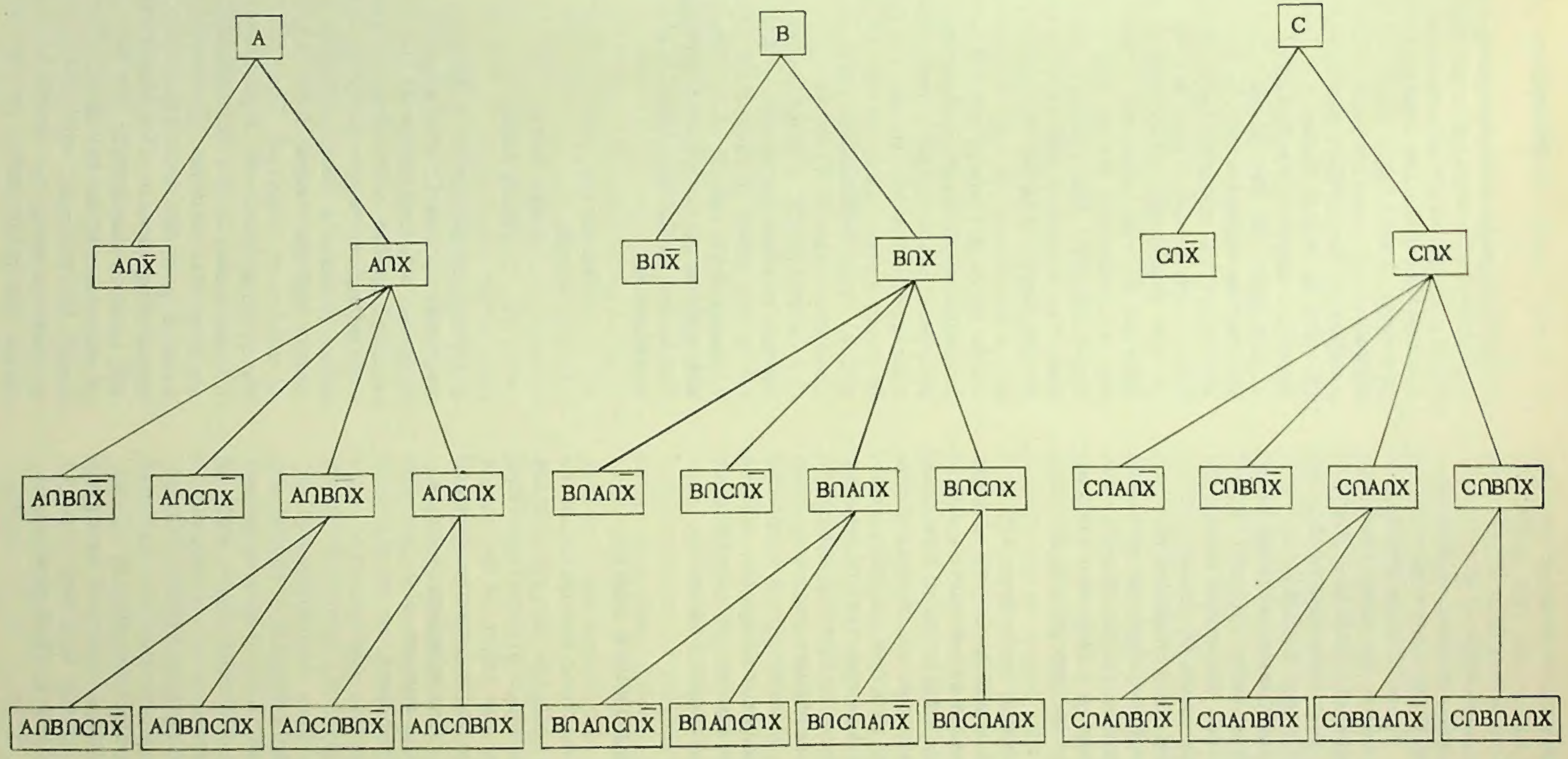
A
B
C
$A\cap\bar{X}$
$A\cap X$
$B\cap\bar{X}$
$B\cap X$
$C\cap\bar{X}$
$C\cap X$
$A\cap B\cap\bar{X}$
$A\cap C\cap\bar{X}$
$A\cap B\cap X$
$A\cap C\cap X$
$B\cap A\cap\bar{X}$
$B\cap C\cap\bar{X}$
$B\cap A\cap X$
$B\cap C\cap X$
$C\cap A\cap\bar{X}$
$C\cap B\cap\bar{X}$
$C\cap A\cap X$
$C\cap B\cap X$
$A\cap B\cap C\cap\bar{X}$
$A\cap B\cap C\cap X$
$A\cap C\cap B\cap\bar{X}$
$A\cap C\cap B\cap X$
$B\cap A\cap C\cap\bar{X}$
$B\cap A\cap C\cap X$
$B\cap C\cap A\cap\bar{X}$
$B\cap C\cap A\cap X$
$C\cap A\cap B\cap\bar{X}$
$C\cap A\cap B\cap X$
$C\cap B\cap A\cap\bar{X}$
$C\cap B\cap A\cap X$

Figure 1.

It will be noted that the same member may be included within the scope of two different classes. Thus $A \cap B \cap \bar{X}$ is a member of $A \cap X$ and also of $B \cap X$. If B is a class defined in terms of the amino characteristic, aniline, which is a member of Class $A \cap B \cap \bar{X}$ is also a member of the benzene class, A, as well as a member of amino class, B. If we follow the diagram down, we can see that a compound which is a member of Class $A \cap B \cap C \cap \bar{X}$ is also a member of Classes A, B, C, $A \cap B \cap X$, $A \cap C \cap X$ and $B \cap C \cap X$.

A search for *a* benzene compound is a generic search for all compounds containing the benzene nucleus and would include the unique compound benzene and compounds such as aniline which contain the benzene ring as a fragment thereof. A search request for anilino compounds excludes benzene but includes aniline and compounds containing the anilino group as a fragment thereof. A search request for aniline specifically includes aniline only ($A \cap B \cap \bar{X}$).

While these illustrations have been with respect to structural characteristics of organic compounds, the principles are similarly applicable to structural characteristics of inorganic compounds and to any type of characteristic such as a use, property, process condition, composition, etc. Thus, if A is the class of benzene compounds, B is a class of starting materials and C is a class of oxidation processes, the combination of these characteristics according to a designated relationship will specify a class containing all benzene compounds involved as starting materials in oxidation processes.

TYPES OF DISCLOSURE UNITS

For convenience in machine manipulation the patent data have been analyzed into elementary bits of disclosure units which are characterized as belonging to one or the other of two major types of subject matter, the substantive type and the relationship type. The substantive type includes, for example, the concept of a chemical element, whether in free form or in chemical combination, the particular manipulative steps and conditions of a physical or chemical process, the description of complex materials and extracts, the functions and behavior of materials. The relationship type pertains to the various associations among the substantive types, such as the way in which the elements are arranged in a structural formula portrayal of a chemical compound, the association of several materials as a physical admixture, the relationship of equivalence of several alternative intermediates for a chemical preparation or the fact that A is related to B as intermediate to final product in a given process.

SCHEDULES

For the purpose of assigning codes to the substantive disclosures, it is necessary to draw up schedules of the structural and functional characteristics which are considered to be pertinent to patent searching. While many of the disadvantages of conventional classification were avoided in the earlier mechanization method by the way in which these scheduled terms were applied, namely, by multiple coding and the elimination of the superiority concept, the limitations encountered by a generic and specific hierarchy of combinations of characteristics still existed. To construct and use a schedule of the magnitude which would be required to provide for all possible combinations of the more elementary terms would be a practical impossibility. Therefore, in the present mechanization system a different method of schedule arrangement is being employed. By this method, all subdivisions of any particular generic or class term pertain to variations of the characteristics of that genus alone and do not involve characteristics of that genus combined with any characteristics pertaining to a different genus. In other words, indented under Class A will not appear any subordinate term which embodies characteristics of Class B.

A schedule of elements may list them by atomic number, by atomic weight, alphabetically or in any other arrangement, and to each will be assigned a unique code. The periodic group will also be coded for each element. A schedule can be set up for the metallic characteristics of the element:

Non-Metal
Metal
    Light Metal
    Heavy Metal.

A term such as "amine" will not have such terms as "with hydroxyl" or "with halogen" indented under it, but such terms as "primary" or "secondary" can be indented under "amine."

In effect, each generic term will form a separate relatively short schedule and disclosure items will be coded from as many of such schedules as required. The relationships among the several codes which pertain to a single item will be shown by devices which will be described later in this article. Each schedule term will have a generic meaning and by means of the modulating device, also described in more detail later in this article, its particular variation of meaning will be indicated. The term "oxidation," for example, will always have the same general meaning, but it will be modulated to indicate its relationship to an item which is an oxidizing agent, a thing which becomes oxidized, and so forth.

Use of the mechanization system by the searcher involves selecting characteristics pertinent to each patent search and relating them into the classes which designate the point of view of the searcher so that selection by machine can be made accordingly. This is complementary to the method of encoding disclosures, which involves coding the various factors from a plurality of viewpoints. Where, for example, a material is disclosed as having structural characteristic A, such as the benzene ring, and function F, such as solvent, both of these aspects of the material are coded, and one of the relationship devices to be described ties the two terms together as plural descriptors of a particular entity, that is, AF refers to a single material having the structure A and the function F. A search for A, regardless of function, would involve the recognition of A within the code AF and would result in retrieving the disclosure of material AF. A search for any material of undesignated structure which has the function F, would also retrieve the disclosure of AF.

RELATIONSHIP DEVICES

The relationships may be handled in various ways. Two of the devices employed at present are "grouping signals" and "interfixes." If a disclosure mentions compounds A, B, C and D and states that the first two are mixed together and the last two are mixed together, the use of a grouping signal might illustrate the situation as (A, B) and (C, D). Here the parentheses would, by a pre-established definition, be known to enclose several things in an admixture relationship, materials in different sets of parentheses therefore being in different mixtures. To illustrate the use of an interfix to describe the same situation we might have: A1 B1 C2 D2. Here, also by a pre-established definition, those compounds having the same numerical attachment are in one mixture and those having different numerical attachments are in different mixtures. The absolute value of the numbers used is of no significance, but their identity or lack of identity is meaningful.

Interfixes can also be used to identify a time sequence relationship. For example, a process may include a series of steps wherein reaction P is followed by reaction Q and this in turn is followed by reaction R. This could be illustrated as P1 Q2 R3. The time sequence may be discovered by applying the defining rule that the larger the interfix number, the later in time is the reaction associated with that number, or stated another way, any reaction is prior in time to one having a larger interfix number. If the process were N1 P1 Q2 R3, it would be apparent that N and P are carried out simultaneously and that both reaction N and reaction P precede Q and R. This might be the case where N and P each results in separate products, these separate products are reacted together in Q, and the material resulting from Q is treated in the R reaction.

Other factors in the reaction, such as temperature, pressure, proportions, manipulative operations and the role of a certain material as catalyst are also encoded and their relationships to the materials involved may be shown by the use of interfixes.

Alternative relationships can be indicated by the use of interfixes. For example, a disclosure may be that compound A is present together with either compound B or compound C, but not with both B and C, i. e., (A) + [(B) or (C)] . The disclosure may then be coded: A1-2 B1 C2. A search for A mixed with B and a search for A mixed with C would each retrieve this disclosure since in each case the two materials sought have the same interfix. However, a search for a mixture of B and C would not retrieve this disclosure since B and C have different interfixes.

Grouping signals may be used to show certain features of the organization of the fragments of a disclosure. As will be shown in more detail, codes for the structures of compounds are based on codes for the elements involved. The codes for the structure of a compound and those for its functions can be associated by grouping signals. Likewise the several compounds making up a mixture can be related to each other by higher order grouping signals, and several compositions disclosed in a process can be grouped together by yet higher order signals. For example,

$$\left\{\left[(AF_A)(BF_B)1\right]\left[(CF_C)(DF_D)2\right]\left[(EF_E)(GF_G)3\right]\right\}$$

would be the plan for the coding of a process in which two compounds having structural characteristics A and B, respectively, are reacted together to form product C, followed by a reaction of C with D to form a mixture of products E and G. F represents the function of each material indicated in the subscript of F. The numerals 1, 2 and 3 are interfixes to indicate the reaction sequences. The parentheses, square brackets and bowed brackets represent successively higher orders of grouping signals. The parentheses group together the plural descriptors of an item, the square brackets group together the items which are simultaneously present as a composition or admixture and the bowed brackets group together the several compositions relating to the same process. The symbols A, B, C, D, E, and G in this example represent all of the codes for the elements contained in each of these compounds and their structural arrangements.

CODING CHEMICAL STRUCTURES

Several different systems for coding structures of compounds have been developed, all based on the same principles but differing in machine manipulation. An attempt has been made to keep the system as universal as possible so that no limitations would be imposed as to the type of machine to be used. However, for purposes of illustration it is necessary to select a particular system, with the understanding that such system may be modified according to the nature of the machine ultimately used to make searches. The structural formula is taken as the basis for coding the structural characteristics of chemical compounds. The method of coding takes cognizance of each element present in the chemical compound and its graphic structural relationship to every other element so that any selected fragment of the entire compound can be recognized and retrieved by machine when a search is made for the class of compounds containing that fragment. In addition, the complete configuration gives a code uniquely different from the code for any other compound.

To illustrate coding using the interfix device, a ring system such as acridine (Fig. 2)

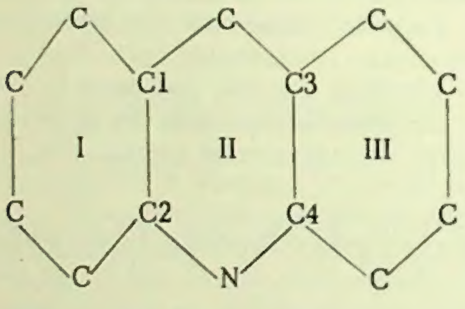

Figure 2.

may be coded as follows:

$$\text{(I)}\quad \underline{\text{r 6/C1 C2 C C C C}}\text{ C1 C2 C C C r}$$

$$\text{(II)}\quad \underbrace{\overbrace{\underline{\text{r 6}}}^{a}}_{b}\underline{\text{/C1 C }}\overbrace{\underline{\text{C3 C4 }}\underbrace{\underline{\text{N C2}}\text{ C1 C C3 C4}}_{b}}^{a}\text{ N }\underbrace{\overbrace{\underline{\text{r}}}^{a}}_{b}$$

$$\text{(III)}\quad \text{r 6/C3 C C C C C4 C3 C C C C r} \quad .$$

The underlines and brackets attached to codes I and II do not relate to the coding of acridine but will be referred to subsequently in the search descriptions. Each element will have a code identification but for ease in visual interpretation the conventional symbols for the elements are used. Each ring is individually coded, starting at any point of the ring and going clockwise, recording the elements in the sequence in which they occur in the ring. "r" represents a code designation for grouping signals indicating a ring and "6" indicates that there are 6 members in the ring. The numerals 1 thru 4 are interfixes which indicate a shared element, e. g., C1 is common to both ring I and ring II.

For reasons that will become apparent the coding sequence is continued for more than a single encirclement of the ring, and the total number of elements coded is calculated as 2 n - 1, where n is the number of elements in the ring.

A search request for a pyridine compound, generically, would be a search for a disclosure of the grouping of Figure 3

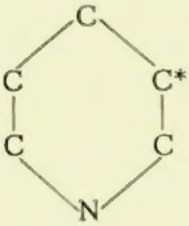

Figure 3.

whether it appears as pyridine per se or as a fragment of a larger molecule such as acridine in Fig. 2. Starting at any point of the ring, for example the carbon marked with the asterisk, the requirement is to find the element sequence r 6/ - - C C N C C C - - r within the stored material containing the coded information. In this search the numerals after the C's in the acridine code are ignored, since they describe how rings are joined, which feature is outside the scope of the search request. It will be seen that the sequence sought appears in the code for ring II (bracketed portions a). If the searcher were to start his sequence at any other place in the ring, e.g., starting at the nitrogen, the sequence would be r 6/ - - N C C C C C - - r which, it will be seen, also appears in the code for ring II of acridine (bracketed portions b). Because of the coding to 2 n - 1, the searcher needs no fixed point of reference for starting to specify the sequence desired, such sequence being picked up at any point in the 2 n - 1 length code.

A search for a quinoline (Fig. 4),

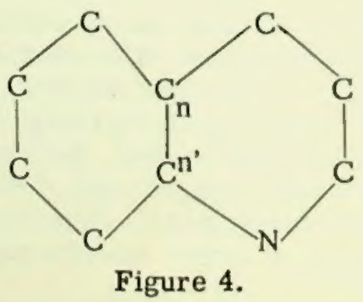

Figure 4.

generically, will be expressed as

r 6/ - - Cn Cn' C C C C - - r

r 6/ - - Cn C C C N Cn' - - r, where "n" is any number and "n' " is any different number. This code will be found within the codes for rings I and II of acridine at the underlined portions. The non-pertinent numerals 3 and 4 are ignored. The requirement is to find the desired pattern *within* the larger pattern, or, in effect, to see whether the requested code subtracts from the stored code. Note that the searcher will not know what interfix numbers were assigned to a particular structure and he can only specify in his search request the sameness or difference of numbers associated with particular elements.

As another example, take the straight chain

$$H - O - CH_2 - CH_2 - O - CH_2 - CH_2Br.$$

In coding, the non-significant hydrogens are left out, as in the ring compounds of the previous examples, i.e., those not needed for identification of a structural group are not coded. The skeleton of the chain will then appear as follows:

H - O - C - C - O - C - C - Br

The code symbols for these elements will be set forth as follows in a sequence which will be handled by the machine in such a way that it may be read from both left to right and right to left: $\underline{\overset{\frown}{Q}\ H - O - C - \overbrace{C - O - C}}$ - C - Br $\underline{\overset{\frown}{Q}}$, where Q represents a code designation for grouping signals indicating a chain.

A search for an ether will be set forth as Q - - C - O - C - - Q, which it will be seen is within the stored code in the bracketed portions.

A search for an ether alcohol in which the two oxygens involved are in a 1 - 4 relationship, will be set forth as Q - - H - O - C - C - O - C - - Q, which is found within the code in the portions underlined.

For a combination of chain and ring, consider the compound (Figure 5)

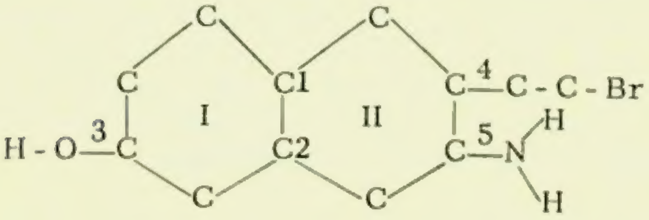

Figure 5.

the code for which is expressed as

(I) r 6/C1 C2 C $C\underset{\cdot}{3}$ C C C1 C2 C $C\underset{\cdot}{3}$ C r

(II) $\underline{r\ 6/}$C1 C $C\underset{\cdot}{4}$ $\underline{C\underset{\cdot}{5}\ C\ C2\ C1\ C\ C\underset{\cdot}{4}}$ $C\underset{\cdot}{5}$ C $\underline{r}$

Q H O $\underset{\cdot}{3}$ Q

Q Br C $C\underset{\cdot}{4}$ Q

$\underline{Q\ H\ N\underset{\cdot}{5}\ H\ Q}$

The structural groups OH, NH and CCBr are treated as straight chains. It is seen that two major types of groups, or what may be termed "building blocks," are recognized, the chains and the rings. Chemical compounds may contain a chain joined to a ring and a ring joined to a ring. A chain joined to a chain is, of course, one chain. The joining of chain to ring is by way of a bond juncture. The joining of ring to ring can be by bond juncture, as in diphenyl, or by element sharing, as in naphthalene. The joining relationship among these "building blocks" is indicated by an interfix number which follows the element of the block wherein this relationship exists. This particular kind of relationship is identified in the Fig. 5 example by the dots beneath numbers 3, 4 and 5 to indicate bond sharing, and the absence of the dots beneath numbers 1 and 2 to indicate element sharing.

A search for *an anilino* compound (Fig. 6)

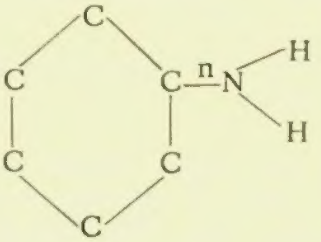

Figure 6.

will be expressed as r 6/ - - $C\underset{\cdot}{n}$ C C C C C - - r

Q H $N\underset{\cdot}{n}$ H Q

This pattern is found in the underlined portions of the Fig. 5 compound code.

Similarly, a search may be made for *a* phenol, for *a* bromoethyl naphthalene, or for *a* naphthol, and so on, and the machine would select disclosures of the compound of Fig. 5 as well as disclosures of all other compounds containing the structural group expressed in the search request, and, therefore belonging to the particular generic class desired. When the search question is so specific as to include only one member within the class, e. g., ethyl alcohol, 2, 5 dinitro-aniline, Peruvian bark, the code sought will be one representing an "index number," which is a unique code representing a specific compound, material or other particular thing.

The index number will be coded as an additional descriptor for the item it represents. Searching by index number will retrieve only disclosures of the specific concept and disclosures generic to the specific thing denoted will not be retrieved.

It can be seen from the preceding examples that the elements in a ring or in a straight chain are coded in the sequence in which they occur in the structure of the ring or straight chain. Branched chain structures are treated in a somewhat different manner because of several complex problems which they present. The methods of coding branched chains will be described in a subsequent article.

MODULANTS

One method mentioned above for indicating some of the contextual relationships of the materials is by the assignment to each material of the function pertaining to it, and modulation of the function to indicate the role of the material with respect to that function.

Modulants may be scheduled as follows:

Potential (active)	- 33
Potential fulfilled (active)	- 33 - 40
Potential (passive)	- 44
Potential fulfilled (passive)	- 44 - 40
Resultant from	- 55

Considering the term "diazotization,'' which is denoted by code 12, modulation of this function would result in the following set of codes:

Diazotization (generic concept)	- 12
Diazotizing agent	- 12 - 33
Diazotizes	- 12 - 33 - 40
Diazotizable	- 12 - 44
Is diazotized	- 12 - 44 - 40
Diazotized compound	- 12 - 55

Each term of function may be thus modified as desired by affixing to it the proper modulant and the resulting code is assigned to each material to which it pertains.

Also, each material in a chemical reaction is assigned a code specifying the process type in which the material is involved, and the process or function code is modulated to indicate whether the compound is a starting material or product produced by the reaction.

A disclosure that A+B diazotize C to make D may be coded as:

(A - SM - 12 - 33 - 40) + (B - SM - 12 - 33 - 40)

+ (C - SM - 12 - 44 - 40) ⟶ (D - FP - 12 - 55)

where SM stands for starting material, FP for final product and A, B, C and D stand for the structural characteristics of compounds A, B, C and D. Note that the verbal function "diazotization" is distributed to each of the compounds related to diazotization.

A search for a diazotization process involving the production of D will be expressed as (D - FP - 12-55). These codes will be found to occur within the above disclosure.

A search for "A" as a diazotizing agent will be:

(A - 12 - 33), which it is noted will select the above disclosure, the potential action or adjective form "Diazotizing agent," being generic to the action or verbal form "Diazotizes," since a disclosure that A diazotizes B indicates that A is a diazotizing agent.

By the device of modulation a variety of meanings may be associated with any particular schedule term. By using the following set of modulants, for example, any term may be modified according to its role in the context of the disclosure.

Disease of	- 50
Disease by	- 60
Infection by	- 60 - 10
Toxicity by	- 60 - 20
Ingredient	- 70

If the term for a particular plant such as a mold is associated with the code 60 - 10, the idea is expressed of a disease by infection with that mold. If the code 50 is associated with the code for the mold, it is thereby indicated that the mold is itself diseased. By association of code 70 with the mold, it is indicated to be present as an ingredient in the composition. The use of code 60 - 20 with a particular material expresses the throught of a disease of poisoning by the indicated material.

The procedures described are still in tentative form and it is expected that many modifications may ensue as the work on mechanization proceeds. Further details will be presented in future articles.

Cooperative efforts and contributions made in connection with the described developments by D. D. Andrews, Director of the Office of Research and Development, and I. Marcus and H. Pfeffer, of the Chemical Staff of the same Office, are gratefully acknowledged.

REFERENCES

(1) "Mechanized Searching in the U.S. Patent Office," M. F. Bailey, B. E. Lanham, and J. Lei-

bowitz, the U.S. Patent Office, Washington 25, D. C. J. Patent Office Society, 35, 566-587, Aug. 1953.

(2) "Report to the Secretary of Commerce by the Advisory Committee on Application of Machines to Patent Office Operations," Washington, D. C., December 22, 1954.

(3) "Patent Office Annual Report, Fiscal Year 1955," also printed in J. Patent Office Society, 38, 106-107, February 1956.

Comm--DC--67039